Matteo Paggiarin 10521399 10:40

Problema 3 :

a) Fittiamo un Linear Model per spiegare la response variabile "Sound level" Y attraverso le covariate:

$$\begin{cases} x : \text{stream frequency} \\ g : \text{air stream velocity} \in \{H, L\} \end{cases}$$

Attraverso l'utilizzo di una dummy variabile per includere la variabile g, stimiamo il modello completo

$$Y = \beta_0(g) + \beta_1(g)\, x + \varepsilon$$

Dal comando lm di R, otteniamo le seguenti stime di β:

$\hat{\beta}_0(H) = 241.9$

$\hat{\beta}_0(L) = 241.9 - 118.5 = 123.36$

$\hat{\beta}_1(H) = 0.0123$

$\hat{\beta}_1(L) = 0.0123 - 0.0025 = 0.00976$

• Diagnostica del test: (guardare fig.1)

Analizziamo la diagnostica dei residui $\hat{\varepsilon}$, attraverso il comando plot(fm):

- eteroschedasticità non presente ✓
- Gaussianità confermata dallo Shapiro test con p-value uguale a 52 % ✓
- assenza di dati outliers con leverage elevato.

↳ possiamo concludere che, data l'omoschedasticità e la Gaussianità di $\hat{\varepsilon}$, l'assunzione $\varepsilon \sim N(0, \sigma^2)$ è rispettata.

Stima di $\hat{\sigma}$: $\boxed{\hat{\sigma} = 43.85.}$

b)

- dipendenza sound level : frequency.

Equivale a testare $H_0 : \begin{cases} \hat{\beta}_1(H) = 0 \\ \hat{\beta}_1(L) = 0 \end{cases}$. R : linear Hypothesis

p. value : 2.043×10^{-11} $\Rightarrow$ rifiutiamo H_0.

C'è dipendenza fra sound level e frequency

- dipendenza sound level : air stream velocity.

Equivale a testare $H_0 : \begin{cases} \hat{\beta}_0(H) - \hat{\beta}_0(L) = 0 \\ \hat{\beta}_1(H) - \hat{\beta}_1(L) = 0 \end{cases}$

↳ in R è sufficiente testare se i coefficienti della dummy variable siano nulli. Usiamo quindi ancora linear Hypothesis : p-value : 7.23×10^{-14}

$\Rightarrow$ rifiutiamo H_0 : Y dipende da $g = \begin{cases} H \\ L \end{cases}$.

- dipendenza sound level : stream velocity a fronte di un aumento unitario di frequenza.

↳ test : $H_0 = \hat{\beta}_1(H) - \hat{\beta}_1(L) = 0$.

Ci viene dato dal summary (fm)!

p-value (velocity L : frequency) = 29 %

↳ ~~accetto~~ non rifiuto H_0!

c) Occorre quindi ridurre il modello a :

$$Y = \beta_0(g) + \beta_1 x + \varepsilon$$

(indipendente da g!

Nuovi parametri:

$\hat{\beta}_0(H) \simeq 255.7$

$\hat{\beta}_0(L) \simeq 111.3$

$\hat{\beta}_1 \simeq 1.08 \times 10^{-2}$

$\hat{\sigma} \simeq 43.92$

d) Stima: $$z_0 = \begin{cases} x = 15\,000 \text{ Hz} \\ g = H \end{cases}$$

Usiamo il comando "predict", e siamo interessati al Confidence Interval. Scegliamo $\alpha = 1\%$.

$$CI_\alpha \left[\mathbb{E}[Y_0 | z_0] \right] \cong [417.48 , 417.78].$$

La stima puntuale di $\mathbb{E}[Y_0 | z_0]$ è di 417.63 decibel.